# Data Science | 30 Days of Machine Learning | Day - 1

Educator Name: Nishant Dhote

Support Team: **+91-7880-113-112**

What is Machine Learning:

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. It evolved from the study of pattern recognition and computational learning theory in artificial intelligence.

Machine learning focuses on the study of algorithms that can learn from data and make predictions (or decisions) based on models built from sample inputs.
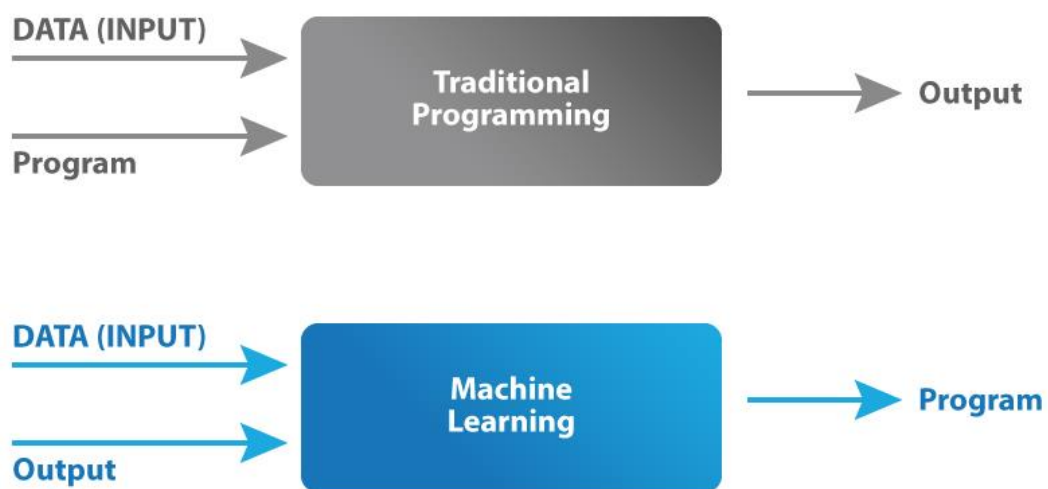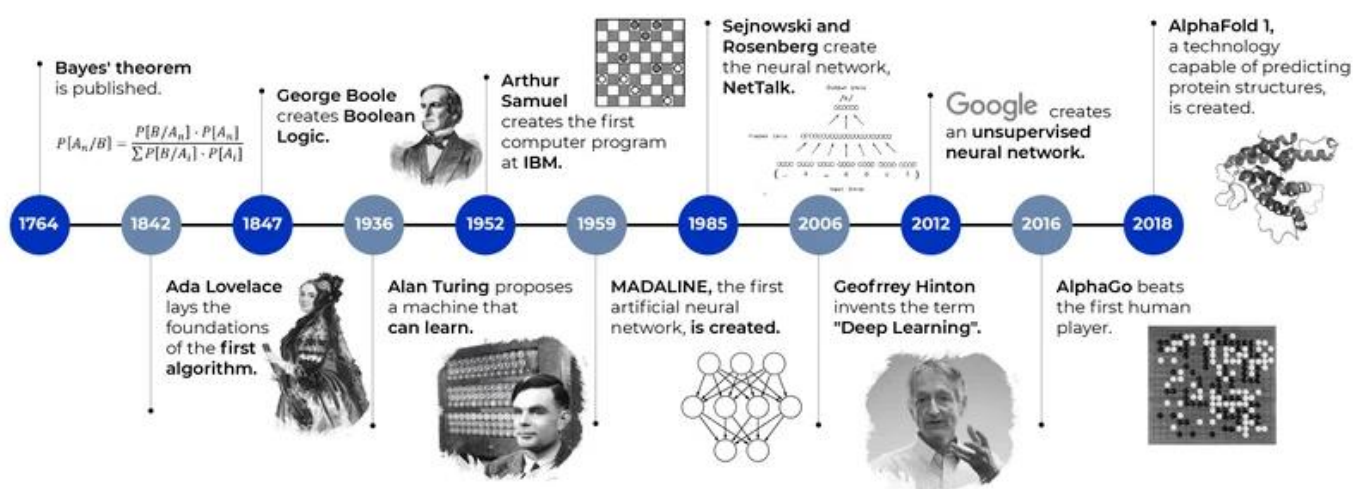

Example: Chess Board Game
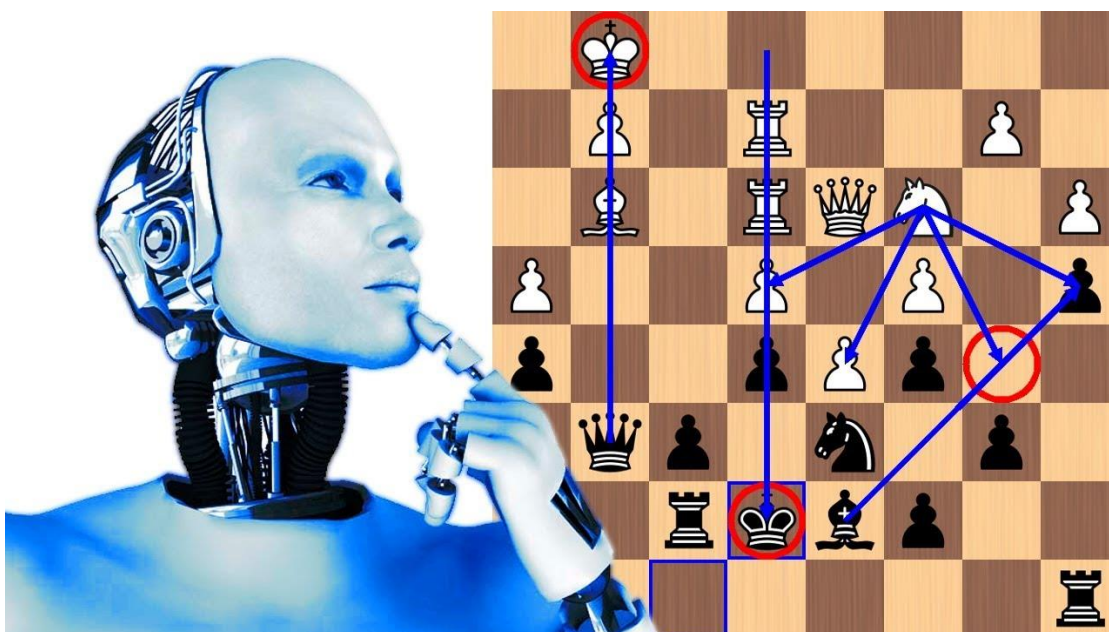
How has Machine Learning evolved? The History of ML

Despite its rise in popularity in the last years, Machine Learning is not a new tool. It has been around for centuries, at least theoretically.

From Thomas Bayes, who in the 18th century laid the foundations of statistics to develop this technology, to the creation of Alpha Go, the first machine to beat a human opponent in the famous game Go, Machine Learning has grown along with humanity.

Here is a brief timeline of the most important events in the history of machine learning
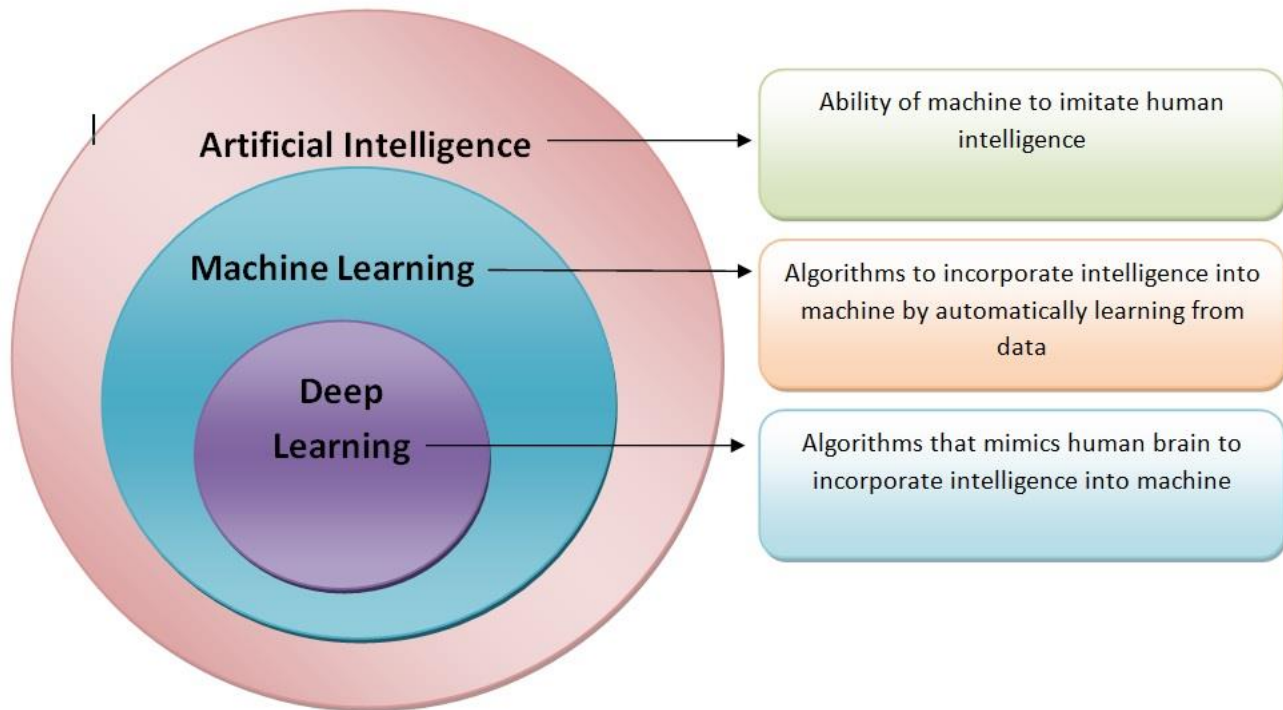
MACHINE LEARNING TIMELINE

Bayes' theorem is published.

$$P[A_n/B] = \frac{P[B/A_n] \cdot P[A_n]}{\sum P[B/A_i] \cdot P[A_i]}$$

George Boole creates **Boolean Logic.**

Arthur Samuel creates the first computer program at IBM.

Sejnowski and Rosenberg create the neural network, **NetTalk.**

Google creates an unsupervised neural network.

AlphaFold 1, a technology capable of predicting protein structures, is created.

1764   1842   1847   1936   1952   1959   1985   2006   2012   2016   2018

Ada Lovelace lays the foundations of the **first algorithm.**

Alan Turing proposes a machine that **can learn.**

MADALINE, the first artificial neural network, **is created.**

Geoffrey Hinton invents the term **"Deep Learning".**

AlphaGo beats the first human player.

Read Article : https://www.techtarget.com/whatis/A-Timeline-of-Machine-Learning-History

Alpha Go: https://deepmind.google/technologies/alphago/

ML vs DL vs AI:

AI simulates human intelligence to perform tasks and make decisions. ML is a subset of AI that uses algorithms to learn patterns from data. DL is a subset of ML that employs artificial neural networks for complex tasks. AI may or may not require large datasets; it can use predefined rules.

Artificial Intelligence → Ability of machine to imitate human intelligence

Machine Learning → Algorithms to incorporate intelligence into machine by automatically learning from data

Deep Learning → Algorithms that mimics human brain to incorporate intelligence into machine

Data Science Vs Data Analytics Vs ML/AI/DL

Data science is an umbrella term that encompasses data analytics, data mining, machine learning, and several other related disciplines. While a data scientist is expected to forecast the future based on past patterns, data analysts extract meaningful insights from various data sources.


artificial intelligence

machine learning

data science

deep learning

big data

data analytics
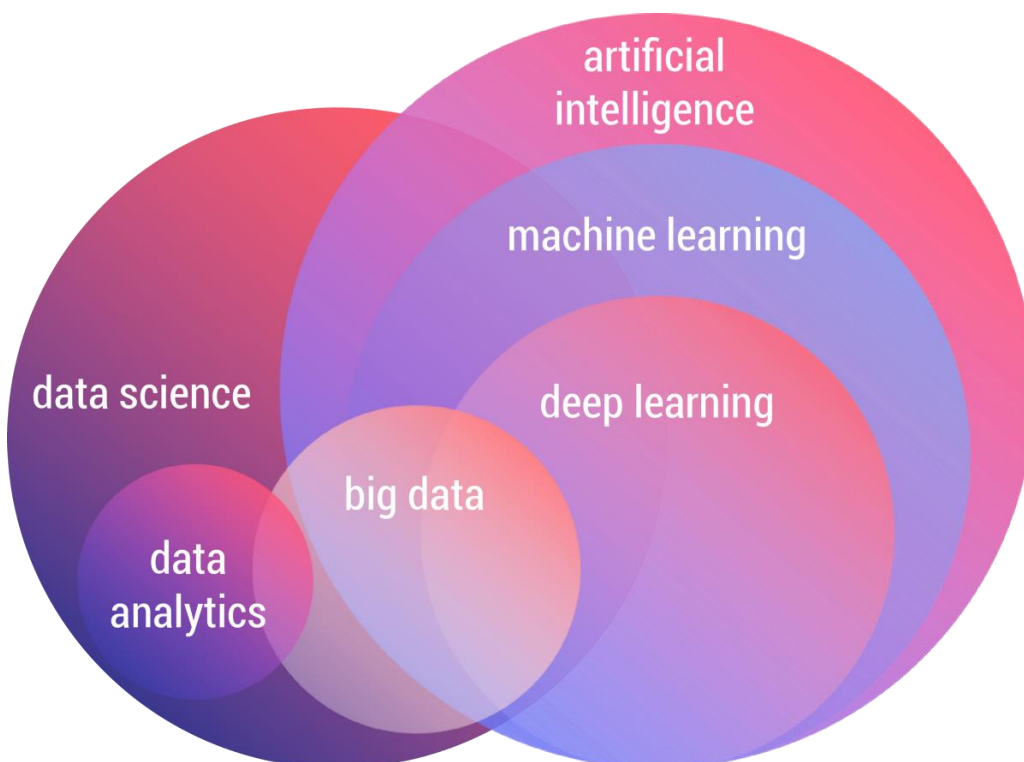
==Types of machine learning==

What are the 4 types of machine learning?

There are four types of machine learning algorithms: supervised, semi-supervised, unsupervised and reinforcement.

**What are the 3 types of machine learning?**

Machine learning involves showing a large volume of data to a machine so that it can learn and make predictions, find patterns, or classify data. The three machine learning types are supervised, unsupervised, and reinforcement learning
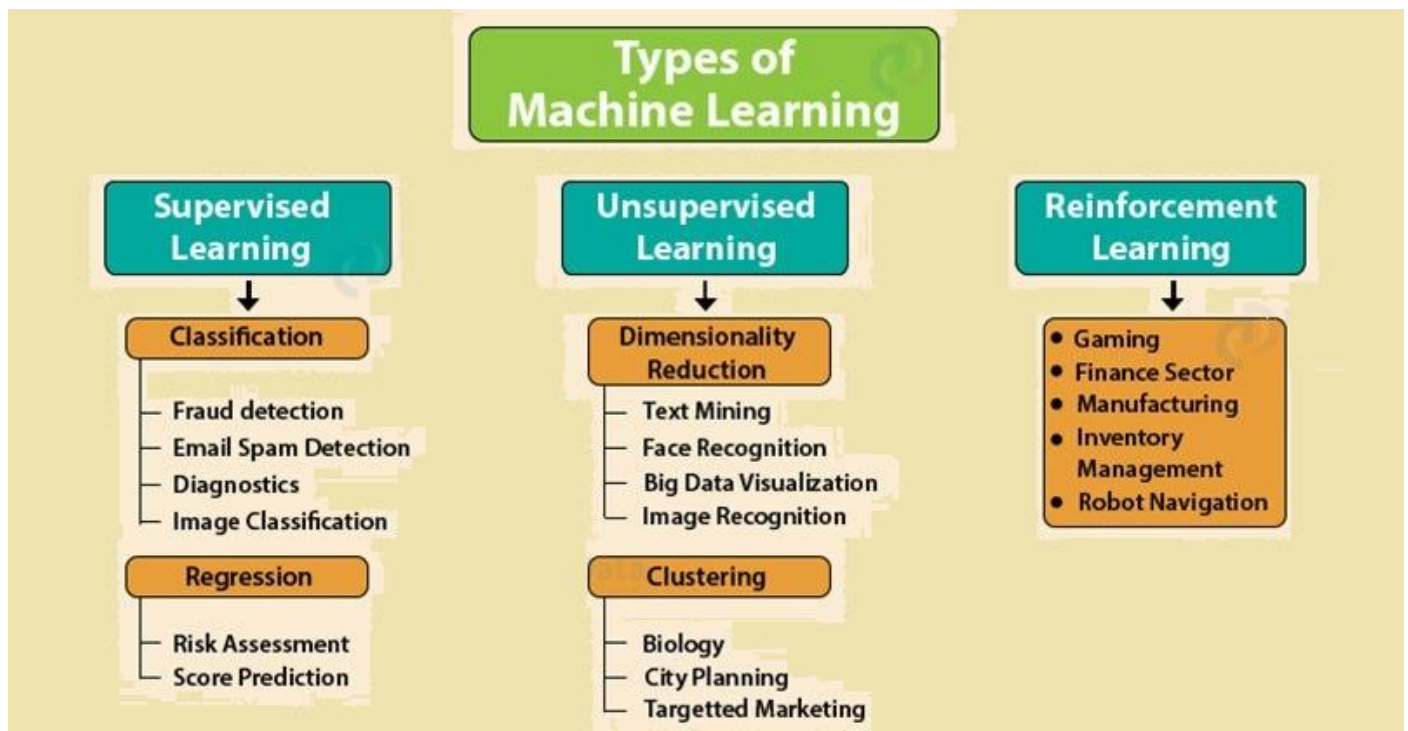
Type of machine learning

1. Supervised ML
   - Regression
   - Classification
2. Un Supervised ML
   - Clustering
   - Dimensionality Reduction
   - Anomaly Detection
   - Association Rule Learning
3. Semi supervised ML
4. Reinforcement ML
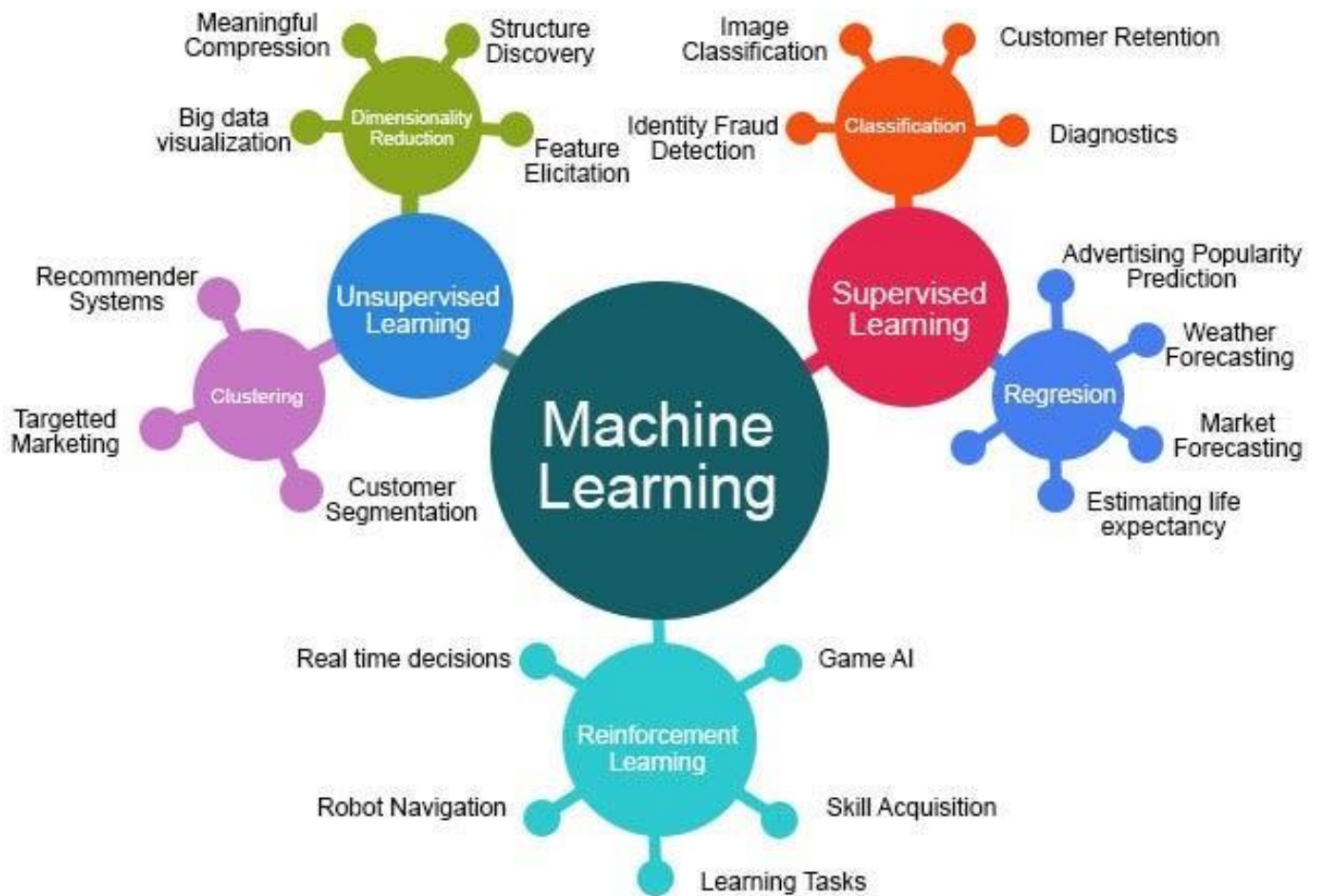
Machine Learning Algorithms

1. Supervised Machine Learning:

Supervised learning is defined as when a model gets trained on a "Labelled Dataset". Labelled datasets have both input and output parameters. In Supervised Learning algorithms learn to map points between inputs and correct outputs. It has both training and validation datasets labelled. Let's understand it with the help of an example.

Example: If we have 10,000 Students Data who have placed or some un placed and we want to find the new predictive result output.

| S.No	Student_Name	St_Percentage	St_CGPA	St_Placement	Package
1	Nishant	87%	7.8	Yes	8 LPA
2	Swati	75%	7.2	No	0 LPA
3	Ragini	91%	8.4	Yes	10 LPA
4	Salman	68%	6.5	No	0 LPA
.					
.					
10K	XYZ	00%	0.0	No	0

Output

Input Data

Conclusion: Draw Mathematical Relationship for generate a new result or indicate the prediction.
Input (87%, 7.8) + Output (Learning ML) = (Program) New Result Output

There are two main categories of supervised learning that are mentioned below:

1. Classification
2. Regression

Data

Numerical Data (%, Age, marks, weight)   Categorical Data (Gender, Name, Nation)
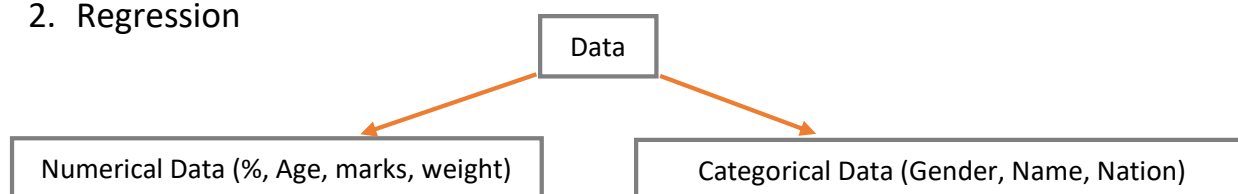
Numerical Output → Regression (Output column is Numerical)
Categorical Output → Classification (Output Column is Categorical)

Example: Consider a scenario where you have to build an image classifier to differentiate between cats and dogs. If you feed the datasets of dogs and cats labelled images to the algorithm, the machine will learn to classify between a dog or a cat from these labelled images. When we input new dog or cat images that it has never seen before, it will use the learned algorithms and predict whether it is a dog or a cat. This is how supervised learning works, and this is particularly an image classification.
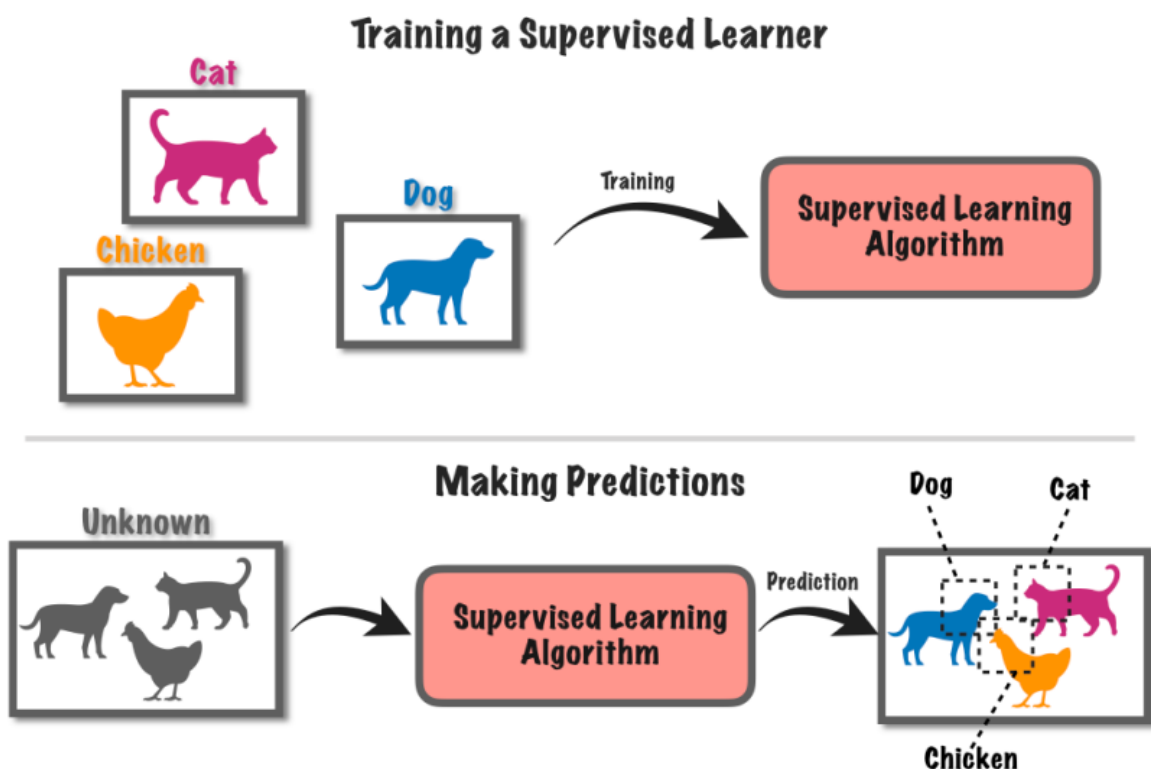
Advantages of Supervised Machine Learning

- Supervised Learning models can have high accuracy as they are trained on labelled data.
- The process of decision-making in supervised learning models is often interpretable.
- It can often be used in pre-trained models which saves time and resources when developing new models from scratch.

Disadvantages of Supervised Machine Learning

- It has limitations in knowing patterns and may struggle with unseen or unexpected patterns that are not present in the training data.
- It can be time-consuming and costly as it relies on labelled data only.
- It may lead to poor generalizations based on new data.

Applications of Supervised Learning

Supervised learning is used in a wide variety of applications, including:

- Image classification: Identify objects, faces, and other features in images.
- Natural language processing: Extract information from text, such as sentiment, entities, and relationships.
- Speech recognition: Convert spoken language into text.
- Recommendation systems: Make personalized recommendations to users.
- Predictive analytics: Predict outcomes, such as sales, customer churn, and stock prices.
- Medical diagnosis: Detect diseases and other medical conditions.
- Fraud detection: Identify fraudulent transactions.
- Autonomous vehicles: Recognize and respond to objects in the environment.
- Email spam detection: Classify emails as spam or not spam.
- Quality control in manufacturing: Inspect products for defects.
- Credit scoring: Assess the risk of a borrower defaulting on a loan.
- Gaming: Recognize characters, analyze player behavior, and create NPCs.
- Customer support: Automate customer support tasks.
- Weather forecasting: Make predictions for temperature, precipitation, and other meteorological parameters.
- Sports analytics: Analyse player performance, make game predictions, and optimize strategies.

2. <mark>Un Supervised Machine Learning:</mark>

Unsupervised learning is a type of machine learning technique in which an algorithm discovers patterns and relationships using unlabelled data. Unlike supervised learning, unsupervised learning doesn't involve providing the algorithm with labelled target outputs. The primary goal of Unsupervised learning is often to discover hidden patterns, similarities, or clusters within the data, which can then be used for various purposes, such as data exploration, visualization, dimensionality reduction, and more.

There are two main categories of unsupervised learning that are mentioned below:
1. Clustering
2. Dimensionality Reduction
(3. Anomaly Detection & 4. Association Rule Learning)

Unsupervised Learning algorithms:

Below is the list of some popular unsupervised learning algorithms.

- K-means clustering
- KNN (k-nearest neighbours)
- Hierarchal clustering
- Anomaly detection
- Neural Networks
- Principle Component Analysis
- Independent Component Analysis
- Singular value decomposition
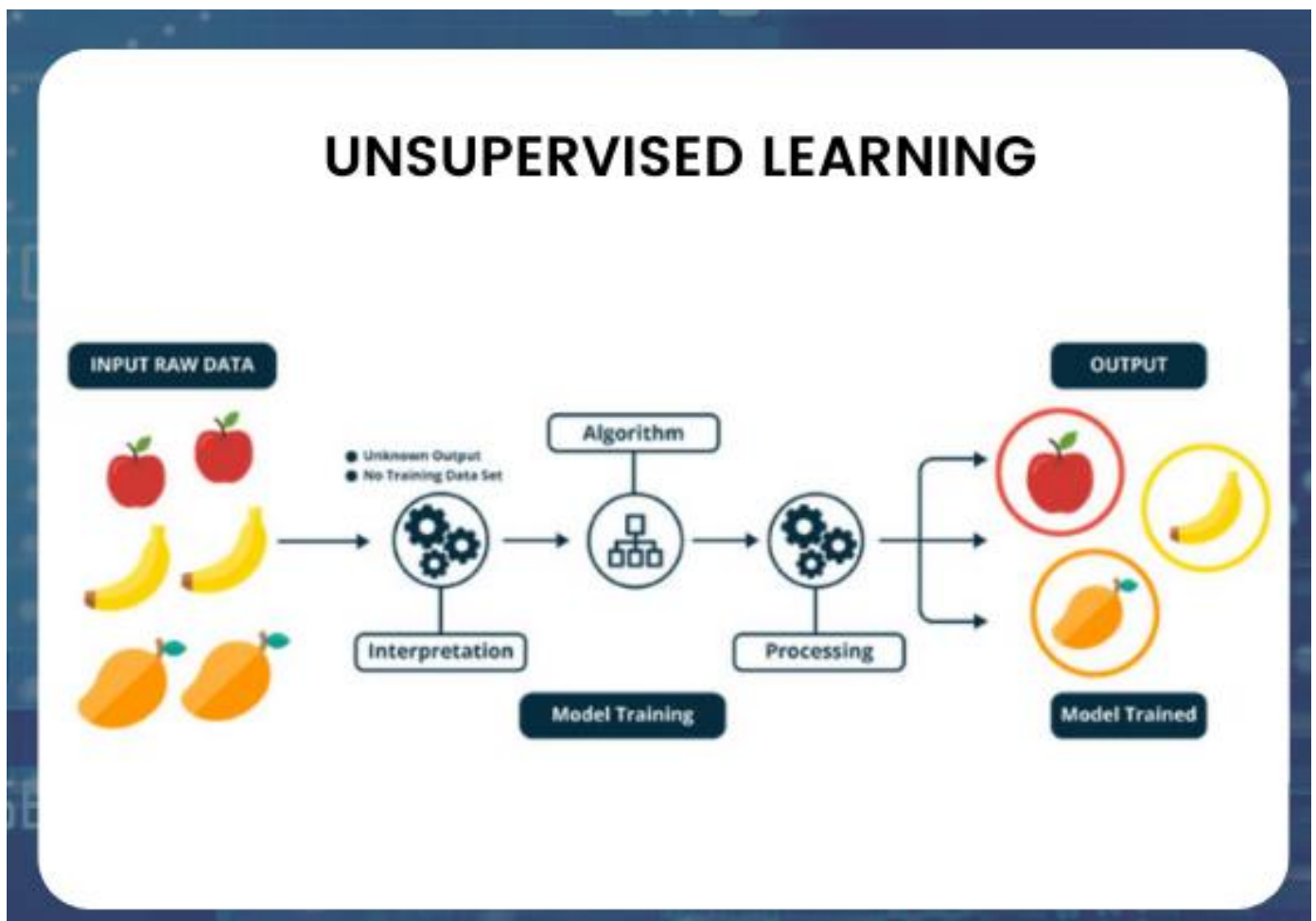
Supervised ML Vs Un-Supervised ML
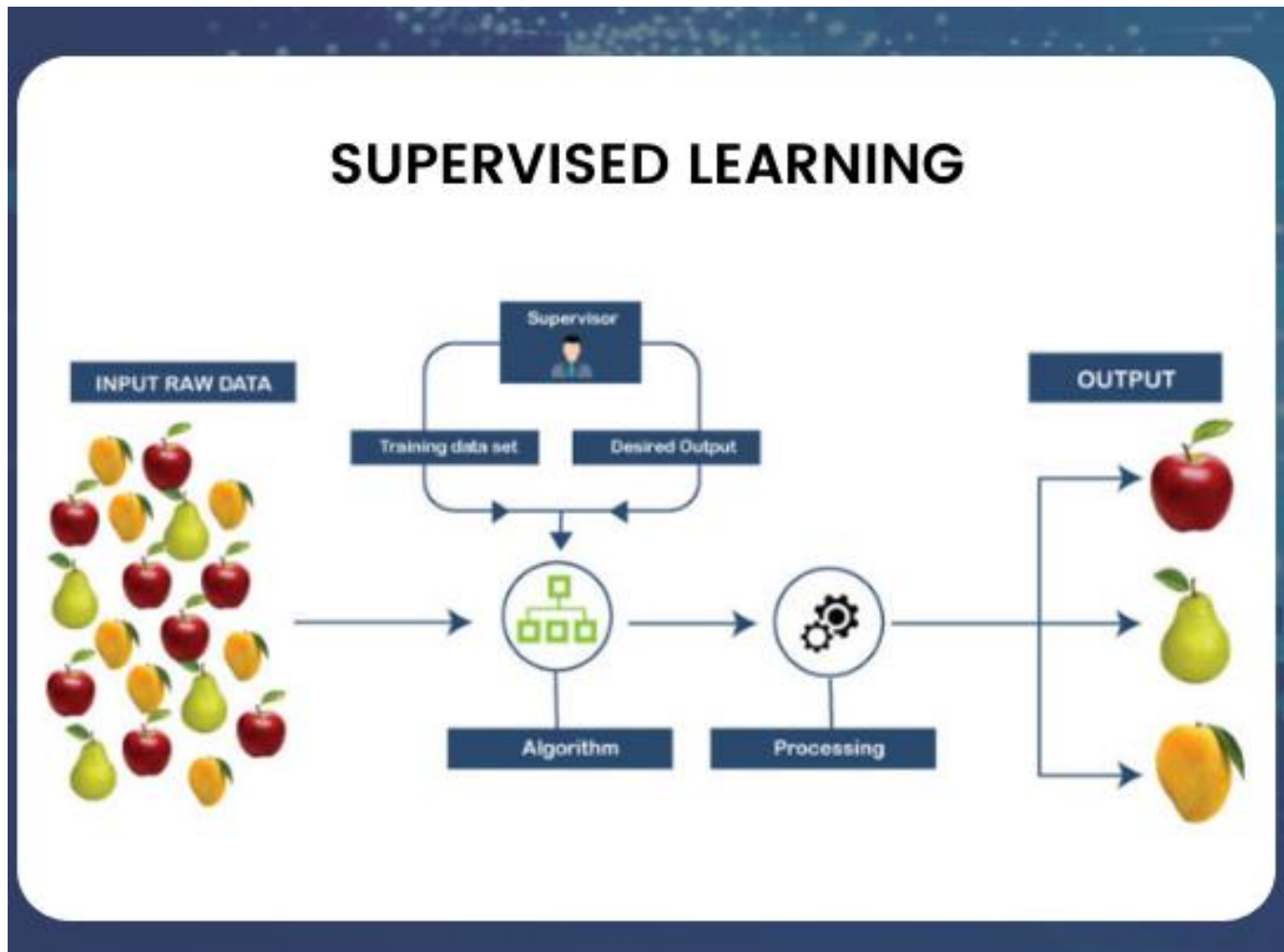
Advantages of Unsupervised Machine Learning

- It helps to discover hidden patterns and various relationships between the data.
- Used for tasks such as customer segmentation, anomaly detection, and data exploration.
- It does not require labelled data and reduces the effort of data labelling.

Disadvantages of Unsupervised Machine Learning

- Without using labels, it may be difficult to predict the quality of the model's output.
- Cluster Interpretability may not be clear and may not have meaningful interpretations.
- It has techniques such as auto encoders and dimensionality reduction that can be used to extract meaningful features from raw data.

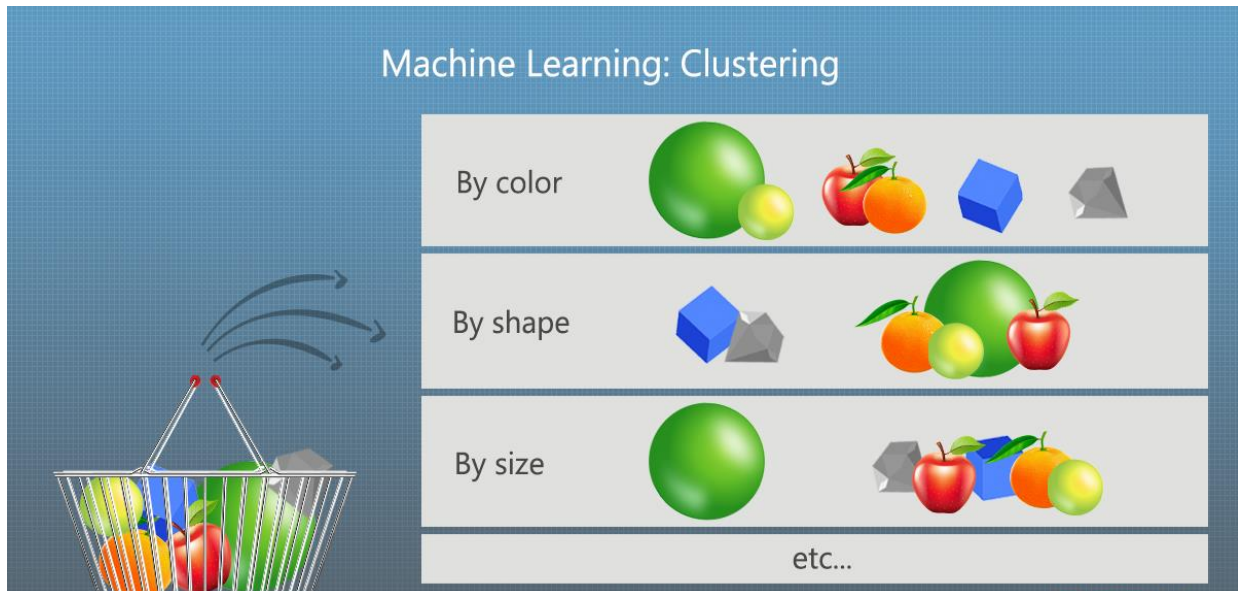## Applications of Unsupervised Learning

Here are some common applications of unsupervised learning:

- Clustering: Group similar data points into clusters.
- Anomaly detection: Identify outliers or anomalies in data.
- Dimensionality reduction: Reduce the dimensionality of data while preserving its essential information.
- Recommendation systems: Suggest products, movies, or content to users based on their historical behaviour or preferences.
- Topic modelling: Discover latent topics within a collection of documents.
- Density estimation: Estimate the probability density function of data.
- Image and video compression: Reduce the amount of storage required for multimedia content.
- Data pre-processing: Help with data pre-processing tasks such as data cleaning, imputation of missing values, and data scaling.
- Market basket analysis: Discover associations between products.
- Genomic data analysis: Identify patterns or group genes with similar expression profiles.
- Image segmentation: Segment images into meaningful regions.
- Community detection in social networks: Identify communities or groups of individuals with similar interests or connections.
- Customer behaviour analysis: Uncover patterns and insights for better marketing and product recommendations.
- Content recommendation: Classify and tag content to make it easier to recommend similar items to users.
- Exploratory data analysis (EDA): Explore data and gain insights before defining specific tasks.

There are 4 main categories of unsupervised learning:
1. Clustering
2. Dimensionality Reduction
3. Anomaly Detection
4. Association Rule Learning

1. Clustering: It is basically a collection of objects on the basis of similarity and dissimilarity between them. Let's understand by example.
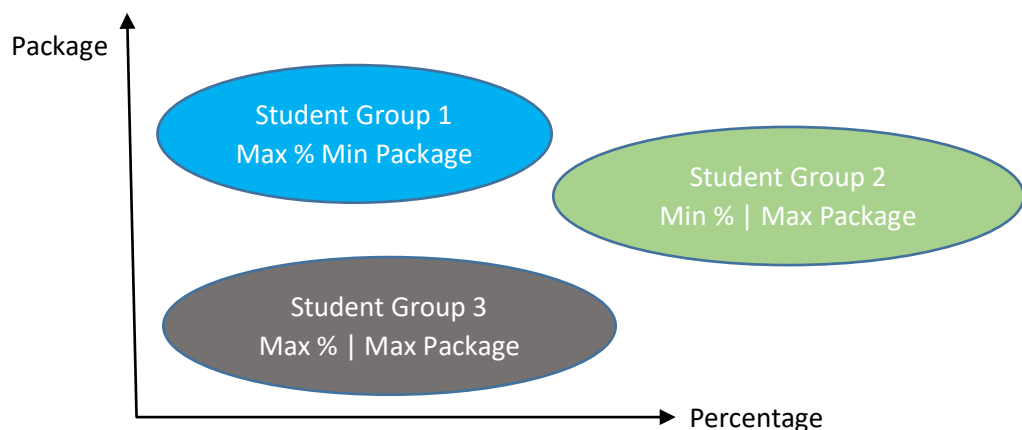

Example: If we have 10,000 Students Data who have placed or some un placed and we want to find the new predictive result output.

Output

S.No	Student_Name	St_Percentage	St_CGPA	St_Placement	Package
1	Nishant	87%	7.8	Yes	8 LPA
2	Swati	75%	7.2	No	0 LPA
3	Ragini	91%	8.4	Yes	10 LPA
4	Salman	68%	6.5	No	0 LPA
.					
.					
10K	XYZ	00%	0.0	No	0

Input Data

2. Dimensionality Reduction: Dimensionality reduction is the process of reducing the number of features (or dimensions) in a dataset while retaining as much information as possible.

Example: Many input columns is not required in the algorithm, it's remove extra columns.

| Type of Rooms	No Of Rooms	Room Size
Bedroom	3	10*15 | 10*10 | 10*12
Washroom	2	6*6 | 6*8
Kitchen	1	10*8

Area
370 Sqft
84 Sqft
80 Sqft

3. Anomaly Detection: Anomaly detection in machine learning is the process of identifying such anomalies in a dataset. Anomaly detection is an important technique in machine learning and data mining, as it can be used to detect unusual behaviour, identify errors, and discover new insights in large datasets.
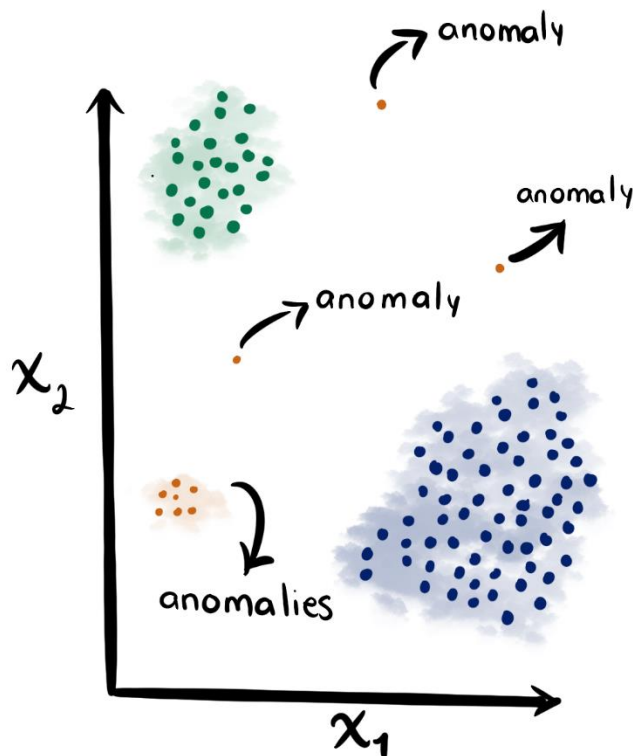
4. Association Rule Learning

Association rule mining is a technique used to uncover hidden relationships between variables in large datasets.
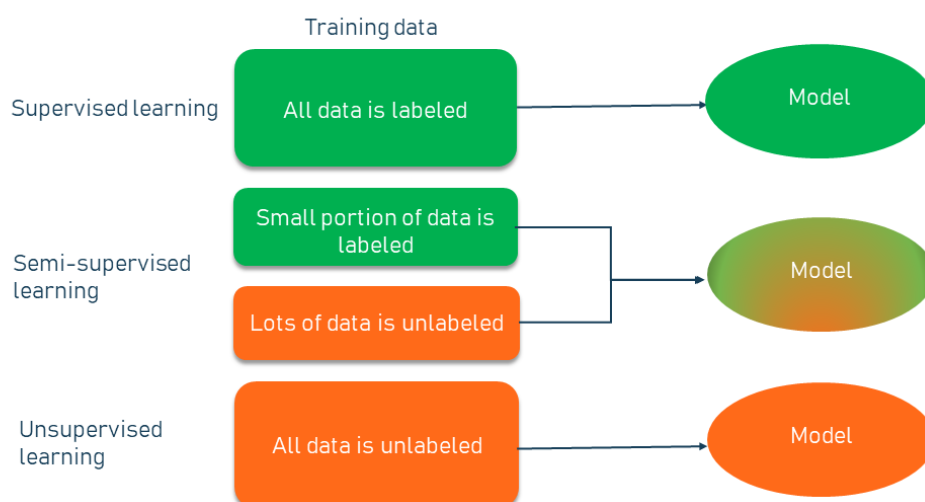
Example: Study many purchase product in shopping mart.
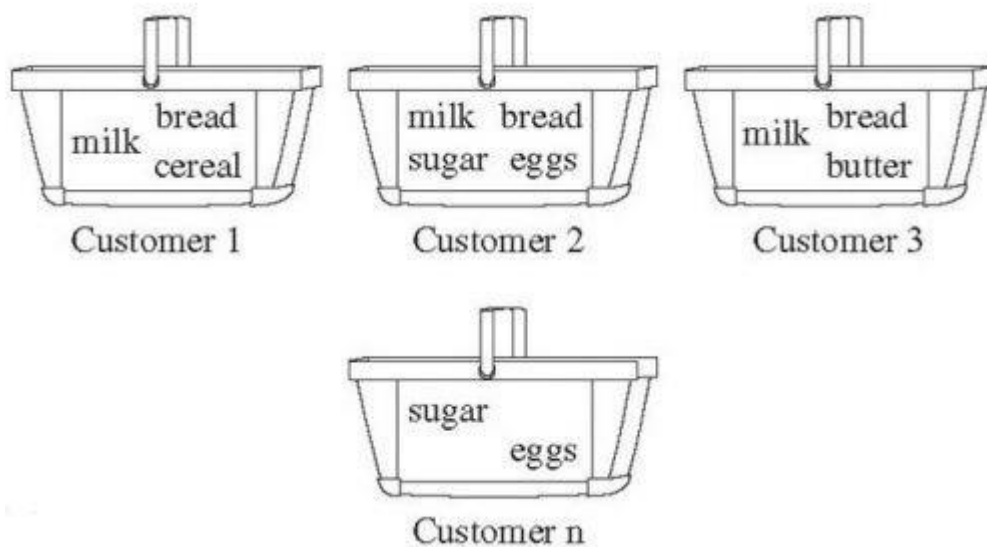
3. **Semi supervised machine learning:** (Partially Supervised and Un Supervised)

We want to ignore labelling, because it's very costly that's why we use "Semi supervised machine learning"

Example: Google photos

SUPERVISED LEARNING vs SEMI-SUPERVISED LEARNING vs UNSUPERVISED LEARNING

4. **Reinforcement machine learning:** Reinforcement learning is a machine learning training method based on rewarding desired behaviours and punishing undesired ones. In general, a reinforcement learning agent -- the entity being trained -- is able to perceive and interpret its environment, take actions and learn through trial and error.

Example: Self Drive Cars | Go Game Deep Mind Company Google